Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly
©2022 ISGSR Organizers. *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_05-012-cd

An Improved Coupled Markov Chain Model for Simulating Geological Uncertainty

Wei Cao[1,2] and Annan Zhou[3]

[1]School of Transportation Science and Engineering, Beihang University, Beijing, China.
[2]School of Architectural Engineering, North China Institute of Science and Technology, Langfang, China
E-mail: by1713111@buaa.edu.cn
[3]School of Engineering, Royal Melbourne Institute of Technology, Melbourne, Australia.
E-mail: annan.zhou@rmit.edu.au

Abstract: Modeling the uncertainty of subsurface stratigraphy is very important for geotechnical engineering problems. Among many models for the geological uncertainty simulation, the coupled Markov chain (CMC) is an effective and widely used model. However, the horizontal transition probability matrix (HTPM) of CMC model is difficult to be determined for geotechnical engineering problems. This paper aims to develop an improved CMC model with an analytical method for estimating the horizontal transition probability matrix (HTPM). The overall tendency of the stratum can be judged by the analytical method and the simulating sequence is determined by the overall tendency. The validity of the overall tendency judgement method is examined through some typical strata. The accuracy of the proposed HTPM estimation method and the strata generated by the improved CMC model are also evaluated. It is found that the overall tendency judgement of each stratum is correct and the overall tendency judgement method is effective. The accuracy of the HTPM estimation and the generated strata are high and insensitive to borehole schemes. The proposed model can well simulate the geological uncertainty based on borehole data.

Keywords: geological uncertainty; coupled Markov chain; horizontal transition probability matrix; analytical method

1 Introduction

It is well acknowledged that the performance of geotechnical constructions is significantly influenced by geological uncertainty (Phoon and Kulhawy, 1999; Ching and Wang, 2015; Li et al., 2016a; Zhou et al., 2016; Gong et al., 2019). Nowadays several methods have been proposed to simulate the strata considering the geological uncertainty. Kohno et al. (1992) introduced a Poisson process to addresses the uncertainties associated with the variability of rock types along the tunnel axis. Elfeki and Dekking (2001; 2005) extended the one-dimensional Markov chain to two dimensions and proposed the coupled Markov chain (CMC) method to model the stratigraphic uncertainty based on borehole data. Bossi et al. (2016) proposed the Boolean Stochastic Generation (BoSG) method to model the soil heterogeneity by stochastically generating different profile configurations. Li et al. (2016b) and Wang et al. (2016) developed a kind of stochastic Markov random field-based approach to characterize the geological uncertainty. Shi and Wang (2021) put forward a non-parametric and data-driven approach to characterize the subsurface stratigraphy based on the multiple point statistics.

Among these models, the CMC model is a simple and useful approach to characterize the geological uncertainty of strata. Many researchers have enhanced this model and applied it to simulate the subsurface stratigraphy (Deng et al., 2018; Li et al., 2019; Qi et al., 2016, 2017; Zhang et al., 2021). However, owing to the wide spacing of boreholes, the horizontal transition probability matrix (HTPM) is difficult to be determined. The existing HTPM estimation methods have some limitations. (1) Using Monte Carlo simulation in the HTPM estimation may lead to a large amount of computation. (2) The simulation sequence is always from left to right which may not be consistent with the overall tendency of strata. In this paper, an improved CMC model is proposed for the geological uncertainty simulation. In this method, the HTPM is estimated using an analytical method and the overall tendency of strata can be judged. Through some typical strata, the effectiveness of the overall tendency judgement, the accuracy of the HTPM estimation and the generated strata using the improved CMC model are evaluated.

2 Improved CMC model

Markov property means that the future depends only on the present and has nothing to do with the past. The CMC model couples two Markov chains in two mutually perpendicular directions. In the CMC model, the vertical transition probability matrix (VTPM) ($\mathbf{P}^{v}$) can be easily estimated from the borehole data. But the HTPM is challenging to be determined because of the sparsity of the known boreholes. To solve this problem, an analytical method for estimating HTPM is established in this paper.

Based on Walther's law, the horizontal transition count matrix (HTCM or $\mathbf{T}^{h}$) and the HTPM ($\mathbf{P}^{h}$) can be calculated using the following formula.

$$T_{ij}^{h}=\begin{cases}T_{ji}^{v} & i\neq j\\ KT_{ij}^{v} & i=j\end{cases} \tag{1}$$

$$p_{ij}^{h}=\frac{T_{ij}^{h}}{\sum_{f=1}^{n}T_{if}^{h}} \tag{2}$$

where T_{ij}^{h}, T_{ij}^{v} and p_{ij}^{h} represent the elements of $\mathbf{T^h}$, $\mathbf{T^v}$ (the vertical transition count matrix) and $\mathbf{P^h}$, respectively. K is an unknown constant representing the ratio of horizontal length to vertical length. Therefore, to estimate the HTPM, the value of K must be determined. The appearance of Borehole i is represented as event A_i (see Figure 1). Assuming the simulating sequence is from left to right, the likelihood of the observed scenario (L) can be expressed as

$$\begin{aligned}L&=\Pr\left(A_2,A_3,\cdots,A_{(N-1)}\middle|A_1,A_N\right)=\Pr(A_2|A_1,A_N)\times\Pr(A_3|A_1,A_2,A_N)\times\cdots\times\Pr\left(A_{(N\text{-}1)}\middle|A_1,A_2,A_3,\cdots,A_{(N\text{-}2)},A_N\right)\\&=\Pr(A_2|A_1,A_N)\times\Pr(A_3|A_2,A_N)\times\cdots\times\Pr(A_{(N\text{-}1)}|A_{(N\text{-}2)},A_N)\end{aligned} \tag{3}$$

Using the properties of coupled Markov chains, the likelihood L (Cao et al., 2021) can be expressed in the following form:

$$L=\prod_{i=2}^{N-1}\left(\frac{p_{S_{C_{i-1},1},S_{C_i,1}}^{h(C_i-C_{i-1})}p_{S_{C_i,1},S_{N_x,1}}^{h(N_x-C_i)}}{p_{S_{C_{i-1},1},S_{N_x,1}}^{h(N_x-C_{i-1})}}\frac{p_{S_{C_{i-1},2},S_{C_i,2}}^{h(C_i-C_{i-1})}p_{S_{C_i,2},S_{N_x,2}}^{h(N_x-C_i)}p_{S_{C_i,1},S_{C_i,2}}^{v}}{\sum_{f=1}^{n}p_{S_{C_{i-1},2},f}^{h(C_i-C_{i-1})}p_{f,S_{N_x,2}}^{h(N_x-C_i)}p_{S_{C_i,1},f}^{v}}\cdots\frac{p_{S_{C_{i-1},N_z},S_{C_i,N_z}}^{h(C_i-C_{i-1})}p_{S_{C_i,N_z},S_{N_x,N_z}}^{h(N_x-C_i)}p_{S_{C_i,1},S_{C_i,N_z}}^{v(N_z-1)}}{\sum_{f=1}^{n}p_{S_{C_{i-1},N_z},f}^{h(C_i-C_{i-1})}p_{f,S_{N_x,N_z}}^{h(N_x-C_i)}p_{S_{C_i,1},f}^{v(N_z-1)}}\right) \tag{4}$$

where p_{ij}^{h} and p_{ij}^{v} represent the elements of $\mathbf{P^h}$ and $\mathbf{P^v}$, respectively. $S_{i,j}$ stands for the soil state at cell (i, j). Based on the maximum likelihood estimation method, considering several K values, the K value corresponding to the maximum value of L is taken as the estimated K_{LR}. Similar to the process above, estimate K_{RL} from right to left, and the estimated K value is the greater one of K_{LR} and K_{RL}. The overall tendency of the stratum and the simulation sequence are associated with the relationship of K_{LR} and K_{RL}. If $K_{\mathrm{LR}}>K_{\mathrm{RL}}$, the overall tendency of the stratum is downward and the simulation sequence is from left to right, otherwise the opposite. If $K_{\mathrm{LR}}=K_{\mathrm{RL}}$, there is no obvious tendency of the stratum, and both of the two simulation sequences are applicable. Then the HTPM can be determined by using Eqs. (1) and (2). Based on the conditional boreholes, the stratum is simulated using the improved CMC model following the simulation sequence associated with the relationship of K_{LR} and K_{RL}. The simulating procedure of the proposed method is detailed in Cao et al. (2021).

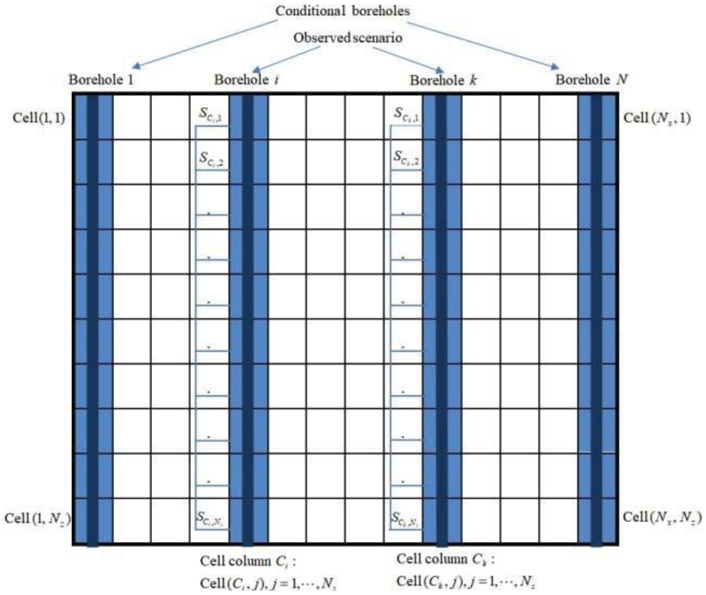

Figure 1. Conditional boreholes and observed scenario.

3 Effectiveness of the overall tendency judgement method

To verify the effectiveness of the overall tendency judgement method, four strata with different types of overall tendency are chosen (see Figure 2). The real overall tendency can be manually recognized based on the real strata. The values of K_{LR} and K_{RL} for different strata are estimated based on the boreholes with the spacing of 3m. Afterwards, the overall tendency of strata is estimated based on the relationship of K_{LR} and K_{RL}. The estimated and real overall tendency of strata are listed in Table 1. It can be seen that the judgements of the overall tendency of strata are all correct which confirmed the validity of the proposed method.

Table 1. Estimated and real overall tendency of different strata.

Stratum ID	K_{LR}	K_{RL}	The relationship between K_{LR} and K_{LR}	Real overall tendency	Estimated overall tendency
Stratum 1	8.8	8.6	$K_{LR}>K_{RL}$	Downward	Downward
Stratum 2	6.4	6.1	$K_{LR}>K_{RL}$	Downward	Downward
Stratum 3	6.0	5.8	$K_{LR}>K_{RL}$	Downward	Downward
Stratum 4	5.5	5.5	$K_{LR}=K_{RL}$	No obvious tendency	No obvious tendency

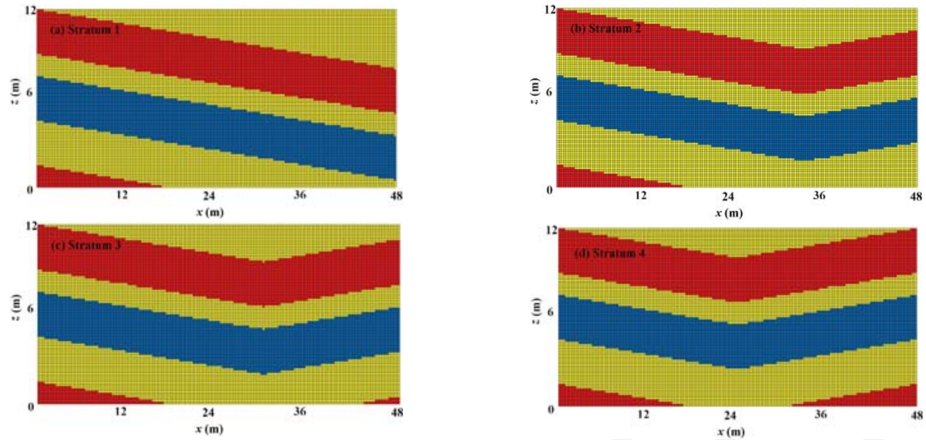

Figure 2. Four real strata with different types of overall tendency (🟥 Clay, 🟨 Sand, and 🟦 Silt).

4 Effectiveness of the improved CMC model

In this section, the effectiveness of the improved CMC model for simulating strata is assessed and the accuracy of the proposed HTPM estimation method is examined through a typical stratum. Figure 3 shows the geological profile with a length of 24m and a depth of 18m. The "real" stratum is an arbitrary realization of CMC model based on the artificial inputs VTPM, HTPM and the two outmost boreholes. The artificial VTPM, HTPM and the real VTPM, HTPM by statistics are listed in Table 2. There are seventeen boreholes including fifteen virtual boreholes and two conditional boreholes in the stratum. Three types of soil, i.e. clay (State 1), sand (State 2) and silt (State 3), are found in the boreholes. The given stratum is divided into cells of 0.3×0.3 m^2 to prepare for the strata simulation.

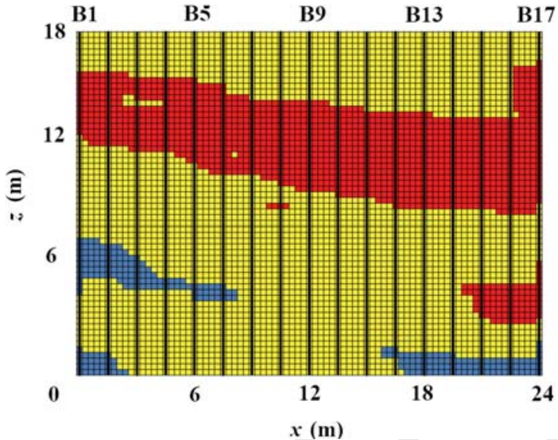

Figure 3. A "real" geologic profile (🟥 Clay, 🟨 Sand, and 🟦 Silt).

Table 2. Artificial and real transition probability matrix of the CMC model.

(a) The artificial VTPM				(b) The artificial HTPM			
State	1	2	3	State	1	2	3
1	0.9348	0.0652	0	1	0.9829	0.0171	0
2	0.0566	0.8868	0.0566	2	0.0156	0.9792	0.0052
3	0	0.0526	0.9474	3	0	0.0400	0.9600
(c) The real VTPM				(d) The real HTPM			
State	1	2	3	State	1	2	3
1	0.9206	0.0794	0	1	0.9886	0.0114	0
2	0.0332	0.9466	0.0202	2	0.0120	0.9843	0.0037
3	0	0.1524	0.8476	3	0	0.0886	0.9114

4.1 Assessment of the HTPM estimation method

In order to study the influence of borehole spacing on the estimation of the HTPM, four schemes (see Table 3) with different borehole spacing (12m, 6m, 3m, 1.5m) are adopted.

Table 3. Four borehole schemes with different borehole spacing.

Borehole scheme	Borehole spacing	Known boreholes
Scheme 1	12m	1, 9, 17
Scheme 2	6m	1, 5, 9, 13, 17
Scheme 3	3m	1, 3, 5, 7, 9, 11, 13, 15, 17
Scheme 4	1.5m	1-17

Tables 4 to 6 show the estimated VTPMs, K values, and HTPMs for all borehole schemes, respectively. For each borehole scheme, the estimated K_{LR} is greater than K_{RL}, so the simulating sequence is from left to right. As shown in Table 6(d), when all the 17 boreholes are known (Scheme 4), the estimated HTPM is very close to the real HTPM. The maximum estimation error of HTPM is |0.9495–0.9114| = 0.0381, which exists in the element p^h_{33}. The estimated HTPMs for different borehole schemes are similar. The difference between the HTPMs for various borehole schemes is small in the elements p^h_{11} and p^h_{33}. The maximum difference is in the element p^h_{22}, and its value is |0.9918–0.9827| = 0.0091. The difference is relatively small in the estimated HTPMs for different borehole schemes, which indicates that the HTPM estimation is not sensitive to borehole spacing. Generally speaking, the estimated HTPMs are very close to the real value for all the borehole schemes, confirming the effectiveness of the proposed method for estimating HTPM.

Table 4. Estimated VTPMs for the four borehole schemes.

(a) Scheme 1				(b) Scheme 2			
State	1	2	3	State	1	2	3
1	0.9355	0.0645	0	1	0.9362	0.0638	0
2	0.0417	0.9271	0.0313	2	0.0341	0.9375	0.0284
3	0	0.0526	0.9474	3	0	0.0800	0.9200
(c) Scheme 3				(d) Scheme 4			
State	1	2	3	State	1	2	3
1	0.9231	0.0764	0	1	0.9233	0.0767	0
2	0.0351	0.9444	0.0205	2	0.0332	0.9456	0.0211
3	0	0.0909	0.9091	3	0	0.1296	0.8704

Table 5. Estimated K values for the four borehole schemes.

Borehole scheme	K_{LR}	K_{RL}	K	Estimated overall tendency	Simulation sequence
Scheme 1	3.2	3.1	3.2	Downward	From left to right
Scheme 2	4.5	4.0	4.5	Downward	From left to right
Scheme 3	5.0	4.4	5.0	Downward	From left to right
Scheme 4	5.6	4.6	5.6	Downward	From left to right

Table 6. Estimated HTPMs for the four borehole schemes.

(a) Scheme 1				(b) Scheme 2			
State	1	2	3	State	1	2	3
1	0.9789	0.0211	0	1	0.9851	0.0149	0
2	0.0138	0.9827	0.0035	2	0.0080	0.9893	0.0027
3	0	0.0495	0.9505	3	0	0.0461	0.9539

(c) Scheme 3				(d) Scheme 4			
State	1	2	3	State	1	2	3
1	0.9836	0.0164	0	1	0.9854	0.0146	0
2	0.0074	0.9908	0.0018	2	0.0062	0.9918	0.0020
3	0	0.0446	0.9554	3	0	0.0505	0.9495

4.2 Strata simulation using the proposed improved CMC model

In terms of the "real" stratum shown in Figure 3, four borehole schemes listed in Table 7 are adopted to study the capacity of the proposed improved CMC model for simulating strata. As shown in Table 7, different borehole schemes have different known boreholes and observation boreholes.

Table 7. Known and observation boreholes of different borehole schemes.

Borehole scheme	Known boreholes	Observation boreholes
Scheme 1	1, 9, 17	3, 5, 7, 11, 13, 15
Scheme 2	1, 5, 9, 17	3, 7, 11, 13, 15
Scheme 3	1, 9, 13, 17	3, 5, 7, 11, 15
Scheme 4	1, 5, 9, 13, 17	3, 7, 11, 15

Two accuracy indexes denoted as I_d and I_w are proposed herein to assess the accuracy of the generated strata from the local and whole perspective, respectively. The index I_d is used to quantify the strata accuracy at the location of observation boreholes, and it is defined as

$$I_d = \frac{1}{N_r}\sum_{i=1}^{N_r} \frac{G_{d,i}}{N_z} \times 100\% \tag{5}$$

where N_r is the times of strata realizations; $G_{d,i}$ is the number of cells matching the soil type of the observation borehole. N_z is the sum of cell rows in the observation borehole. The index I_w is used to quantify the whole strata accuracy, and it is defined as

$$I_w = \frac{\sum_{x=1}^{N_x}\sum_{z=1}^{N_z} I(x,z)}{N_x \times N_z} \times 100\% \tag{6}$$

$$I(x,z) = \begin{cases} 1, Z(x,z) = Z_R(x,z) \\ 0, Z(x,z) \neq Z_R(x,z) \end{cases} \tag{7}$$

where N_x and N_z are the sums of cell rows and columns, respectively; $Z(x,z)$ and $Z_R(x,z)$ denote the soil type at cell (*x*, *z*) of the simulated stratum and the real stratum, respectively.

For each scheme, 1000 strata are simulated using the improved CMC model and Figure 4 shows typical stratigraphic realizations for different borehole schemes. Then the two accuracy indexes are calculated. As shown in Table 8, the values of I_d are very high and all of them are greater than 77.61%. From Scheme 1 to Scheme 4, the known borehole number is increasing, and the accuracy of most observation boreholes is improved. Through the comparison of scheme 2 and scheme 1, the accuracy index values located in B3 and B7 increase from 78.50% and 88.64% to 81.95% and 94.46%, respectively, while the accuracy index values located in B11, B13 and B15 do not increase obviously. This shows that additional borehole can improve the accuracy of the adjacent area, but have little influence on the accuracy of the remote area. As shown in Figure 5, the average values of I_w for the four schemes are all greater than 85.61%. With the increase of borehole number, the average value of I_w increases and the standard deviation of I_w decreases, which indicates that the simulated strata are closer to the real stratum when more boreholes are known. In general, the improved CMC model performs well in simulating the strata for all the borehole schemes.

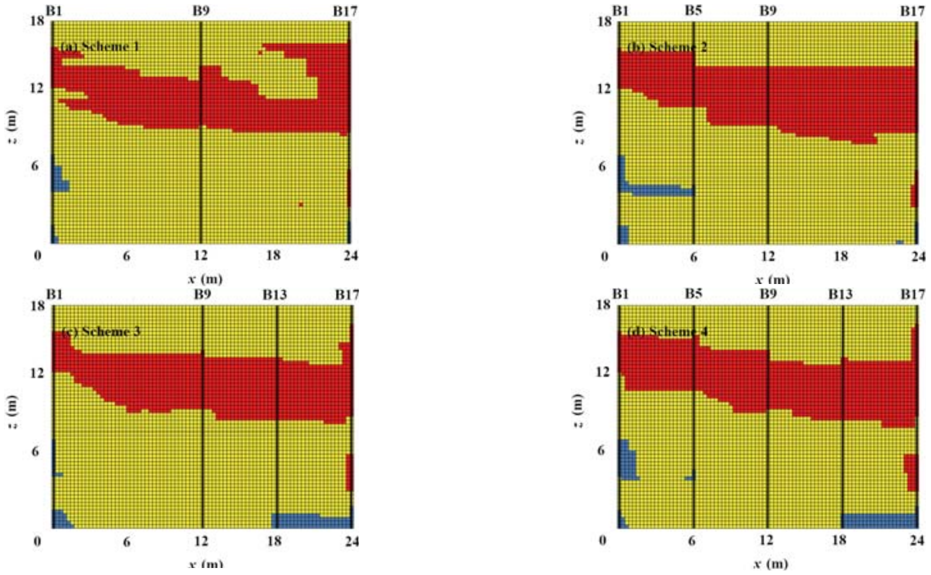

Figure 4. Typical stratigraphic realizations for different borehole schemes.

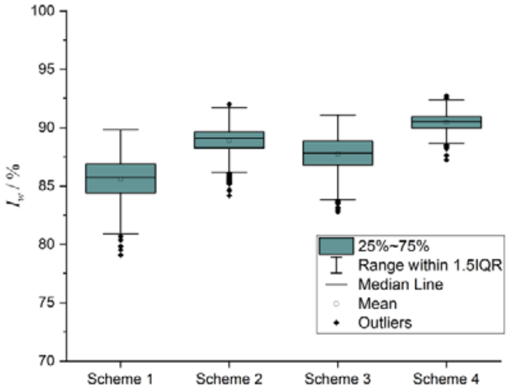

Figure 5. Boxplots of accuracy index I_w for different borehole schemes.

Table 8. Accuracy index I_d for different borehole schemes.

I_d (%)	B1	B3	B5	B7	B9	B11	B13	B15	B17
Scheme 1	—	78.50	77.61	88.64	—	94.36	86.16	79.21	—
Scheme 2	—	81.95	—	94.46	—	95.28	87.99	81.05	—
Scheme 3	—	78.54	77.81	89.22	—	94.93	—	86.17	—
Scheme 4	—	81.95	—	94.44	—	95.49	—	86.60	—

5 Conclusions

In this paper, an improved CMC model with an analytical method for HTPM estimation is established for simulating geological uncertainty. Several strata are chosen to evaluate the effectiveness of the proposed model. It is found that the overall tendency judgement method is effective for all the strata. The accuracy of the estimated HTPMs is high and insensitive to borehole schemes. The proposed two indexes are high for all the borehole schemes, which indicates that the proposed model can well simulate the strata both from the local and whole perspective based on borehole data.

Acknowledgments

The authors appreciate the financial support of the National Natural Science Foundation of China (Project Nos. 42177170, 52004090), the Natural Science Foundation of Hebei Province (E2021508031) and Langfang Science and Technology Research and Development Program (No. 2020013026).

References

Bossi, G., Borgatti, L., Gottardi, G., Marcato, G. (2016). The Boolean Stochastic Generation method - BoSG: a tool for the analysis of the error associated with the simplification of the stratigraphy in geotechnical models. *Engineering Geology*, 203, 99-106.

Cao, W., Zhou A., Shen S.L. (2021). An analytical method for estimating horizontal transition probability matrix of coupled Markov chain for simulating geological uncertainty. *Computers and Geotechnics*, 129.

Ching, J., Wang, J.S. (2015). Application of the transitional Markov chain Monte Carlo algorithm to probabilistic site characterization. *Engineering Geology*, 203, 151-167.

Deng, Z.P., Li, D.Q., Qi, X.H., Cao, Z.J. (2018). Simulation of geological uncertainty using modified generalized coupled Markov chain. *Chinese Journal of Geotechnical Engineering*, 40(11), 2041-2050.

Elfeki, A., Dekking, M. (2001). A Markov chain model for subsurface characterization: Theory and applications. *Mathematical Geology*, 33(5), 569-589.

Elfeki, A.M.M., Dekking, F.M. (2005). Modelling subsurface heterogeneity by coupled Markov chains: Directional dependency, Walther's law and entropy. *Geotechnical and Geological Engineering*, 23(6), 721-756.

Gong W., Tang H., Wang H., Wang X., Juang, C.H. (2019). Probabilistic analysis and design of stabilizing piles in slope considering stratigraphic uncertainty. *Engineering Geology*, 259. 105162.

Li, D.Q., Qi, X.H., Cao, Z.J., Tang, X.S., Phoon, K.K., Zhou, C.B. (2016a). Evaluating slope stability uncertainty using coupled Markov chain. *Computers and Geotechnics*, 73, 72-82.

Li, J., Cai, Y., Li, X., Zhang, L. (2019). Simulating realistic geological stratigraphy using direction-dependent coupled Markov chain model. *Computers and Geotechnics*, 115.

Li, Z., Wang, X., Wang, H., Liang, R.Y. (2016b). Quantifying stratigraphic uncertainties by stochastic simulation techniques based on Markov random field. *Engineering Geology*, 201, 106-122.

Kohno, S., Ang, H. S., Tang, W. H. (1992). Reliability evaluation of idealized tunnel systems. *Structural Safety*, 11(2), 81-93.

Phoon, K.K., Kulhawy, F.H. (1999). Characterization of geotechnical variability. *Canadian Geotechnical Journal*, 36(4), 612-624.

Qi, X.H., Li, D.Q., Phoon, K.K., Cao, Z.J., Tang, X.S. (2016). Simulation of geologic uncertainty using coupled Markov chain. *Engineering Geology*, 207, 129-140.

Qi, X.H., Li, D.Q., Zhou, C., Phoon, K.K. (2017). Estimation of Horizontal Transition Probability Matrix for Coupled Markov Chain Based on Borehole Data. *Journal of Basic Science and Engineering*, 25(5), 967-984.

Shi, C., Wang, Y. (2021). Nonparametric and data-driven interpolation of subsurface soil stratigraphy from limited data using multiple point statistics. *Canadian Geotechnical Journal*, 58, 261-280.

Wang, X., Li, Z., Wang, H., Rong, Q., Liang, R.Y. (2016). Probabilistic analysis of shield-driven tunnel in multiple strata considering stratigraphic uncertainty. *Structural Safety*, 62, 88-100.

Zhang, J.Z., Huang, H.W., Zhang, D.M., Phoon, K.K., Liu, Z.Q., Tang, C. (2021). Quantitative evaluation of geological uncertainty and its influence on tunnel structural performance using improved coupled Markov chain. *Acta Geotechnica*.

Zhou, A., Li, C.Q., Huang, J. (2016). Failure analysis of an infinite unsaturated soil slope. *Proceedings of the Institution of Civil Engineers Geotechnical Engineering*, 169(5), 410-420.